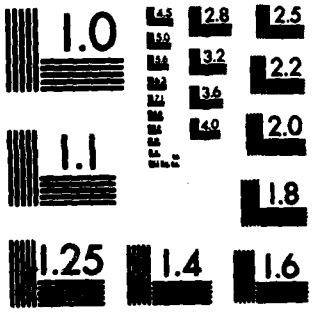
AD-A122 653 SOME OBSERVATIONS ON THE GENERALIZED CONJUGATE GRADIENT 1/1
METHOD(U) YALE UNIV NEW HAVEN CT DEPT OF COMPUTER
SCIENCE S C EISENSTAT SEP 82 RR-244 N00014-76-C-0277

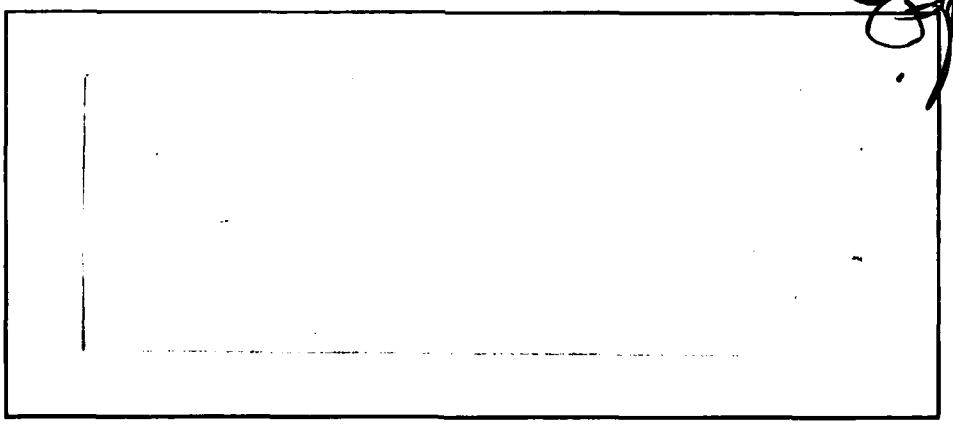
UNCLASSIFIED F/G 12/1 NL

END
DATE FILMED
2 83
DTIC



(12)

AD A122653



YALE UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

(12)

Some Observations on the Generalized Conjugate Gradient Method[†]

Stanley C. Eisenstat[‡]

Research Report #244
September 1982

[†] Presented at the Interamerican Workshop on Numerical Methods, Caracas, Venezuela, June 14-18, 1982.

[‡] Department of Computer Science and Research Center for Scientific Computation, Yale University, P. O. Box 2158, New Haven, CT 06520.

The work presented in this paper was supported in part by the Office of Naval Research under contracts N00014-76-C-0277 and N00014-82-K-0184 and by the National Science Foundation under grant MCS-81-04874.

Some Observations on the Generalized Conjugate Gradient Method

Stanley C. Eisenstat

DEPARTMENT OF COMPUTER SCIENCE AND
RESEARCH CENTER FOR SCIENTIFIC COMPUTATION
YALE UNIVERSITY, NEW HAVEN, CT 06520/USA

1. Introduction

Consider the system of linear equations

$$Ax = b \tag{1}$$

where the coefficient matrix A is large and sparse and has positive definite symmetric part $M = (A+A^t)/2$. In this paper, we compare two methods for solving such systems, the generalized conjugate gradient method of Concus and Golub [2] and Widlund [10] and Craig's method (see [6]) applied to a symmetrically preconditioned auxiliary system.

Notation: (y,z) denotes the Euclidean inner-product $y^t z$ and $\|\cdot\|$ the corresponding norm. If Q is a symmetric positive definite matrix, then $(y,z)_Q$ denotes the Q-inner product (Qy,z) and $\|\cdot\|_Q$ the corresponding norm; $Q^{1/2}$ denotes any square root of Q; and $Q^{-1/2}$ denotes $[Q^{1/2}]^{-1}$. Let $A = M-N$, whence $-N = (A-A^t)/2$ is the skew-symmetric part of A; let $K = M^{-1}N$; and let $\Lambda = \|K\|_M$.

2. The Generalized Conjugate Gradient Method

Concus and Golub [2] and Widlund [10] proposed the Generalized Conjugate Gradient (GCG) method for use when systems of the form $Mz = d$ are "easy" to solve (much more so than the original system):

LET $x^{(0)}$ BE GIVEN AND SET $x^{(-1)} = 0$.

FOR $m = 0$ STEP 1 UNTIL "CONVERGENCE" DO

$r^{(m)} = b - Ax^{(m)}$

SOLVE $Mv^{(m)} = r^{(m)}$

$\rho_m = (r^{(m)}, v^{(m)})$

$$\omega_{m+1} = \begin{cases} 1 & m = 0 \\ [1 + \rho_m/(\rho_{m-1}\omega_m)]^{-1} & m > 0 \end{cases}$$

$$x^{(m+1)} = x^{(m-1)} + \omega_{m+1}\left[v^{(m)} + x^{(m)} - x^{(m-1)}\right]$$

The cost per iteration is one matrix multiply (by A), one solve of a system of the form $Mz = d$, and $2n$ multiplies.

The iterate $x^{(m)}$ can be characterized as the unique element in the affine Krylov subspace

$$x^{(0)} + \text{Span}\{v^{(0)}, Kv^{(0)}, K^2v^{(0)}, \ldots, K^{m-1}v^{(0)}\} \equiv x^{(0)} + S_m$$

satisfying the Galerkin condition

$$(z, Ax^{(m)} - b) = 0 \qquad \text{for all } z \in S_m$$

(see [10]). Moreover, it can be shown that

$$x^{(m)} = x + \pi_m(K)(x^{(0)} - x) \ , \tag{2}$$

where $\pi_m(\mu)$ is an even (odd) polynomial of degree at most m for m even (odd) and $\pi_m(1) = 1$ (see [10]).

The iterate $x^{(m)}$ can also be characterized as the best approximation from a certain m-dimensional affine subspace:[1]

$$x^{(m)} = \begin{cases} \underset{y \in x^{(0)} + (I+K)S_{2k}}{\arg\min} \|y - x\|_M & m \text{ even } (= 2k) \\ \underset{y \in x^{(0)} + v^{(0)} + (I+K)S_{2k+1}}{\arg\min} \|y - x\|_M & m \text{ odd } (= 2k+1) \end{cases}$$

(see [4]). Equivalently,

$$\|x^{(m)} - x\|_M \leq \|p_m(K)(x^{(0)} - x)\|_M \qquad \text{for all } p_m \in P_m,\ 1$$

where P_m is the space of all real polynomials $p_m(\mu)$ of degree at most m satisfying $p_m(1) = 1$ and $p_m(-1) = (-1)^m$ (see [4]). Taking $p_m(\mu) = T_m(i\Lambda^{-1}\mu)/T_m(i\Lambda^{-1})$, where $T_m(z)$ is the m^{th} Chebyshev polynomial, yields the error bound[2]

$$\|x^{(m)} - x\|_M \leq \frac{2}{R(\Lambda)^m + [-R(\Lambda)]^{-m}} \|x^{(0)} - x\|_M \ , \tag{3}$$

where $R(\Lambda) = \Lambda^{-1} + \sqrt{\Lambda^{-2}+1}$. Taking $p_m(\mu) = \mu\pi_{m-1}(\mu)$ yields the inequality

---

[1] However, $x^{(m)}$ is *not* the best approximation to x from the "natural" affine Krylov subspace $x^{(0)} + S_m$ (see [4]).

[2] The case m even was first proved by Hageman, Luk, and Young [8]; Widlund [16] gives a somewhat weaker bound.

$$\Lambda^{-1}\,\|x^{(m+1)}-x\|_M \;\le\; \|x^{(m)}-x\|_M \;\le\; \Lambda\,\|x^{(m-1)}-x\|_M \qquad \text{for all } m \ge 1, \tag{4}$$

which shows that *the even and odd iterates must exhibit the same asymptotic rate of convergence* (cf. [10]).

3. The CSP Method

For any symmetric positive definite matrix Q, the system (1) is equivalent to the symmetrically preconditioned system

$$\hat{A}\hat{x} \;\equiv\; [Q^{-1/2}AQ^{-1/2}]\,[Q^{1/2}x] \;=\; [Q^{-1/2}b] \;\equiv\; \hat{b}\ .$$

If we apply Craig's method (see [6]) to this auxiliary system, which is equivalent to applying the conjugate gradient method to the normal equations

$$\hat{A}\hat{A}^t\hat{y} = \hat{b}\ , \qquad \hat{x} = \hat{A}^t\hat{y}\ ,$$

then the resulting method, Craig's method applied to the Symmetrically Preconditioned auxiliary system (CSP), can be expressed directly in terms of A, x, b, and Q (see [5]):

LET $y^{(0)}$ ($\equiv x^{(0)}$) BE GIVEN

$r^{(0)} = b - Ay^{(0)}$

SOLVE $Q\bar{r}^{(0)} = r^{(0)}$

$\bar{p}^{(0)} = A^t\bar{r}^{(0)}$

SOLVE $Qp^{(0)} = \bar{p}^{(0)}$

FOR $k = 0$ STEP 1 UNTIL "CONVERGENCE" DO

$$\alpha_k = \frac{(r^{(k)},\,\bar{r}^{(k)})}{(p^{(k)},\,\bar{p}^{(k)})}$$

$y^{(k+1)} = y^{(k)} + \alpha_k p^{(k)}$

$r^{(k+1)} = r^{(k)} - \alpha_k Ap^{(k)}$

SOLVE $Q\bar{r}^{(k)} = r^{(k)}$

$$\beta_k = \frac{(r^{(k+1)},\,\bar{r}^{(k+1)})}{(r^{(k)},\,\bar{r}^{(k)})}$$

$\bar{p}^{(k+1)} = A^t\bar{r}^{(k+1)} + \beta_k\bar{p}^{(k)}$

SOLVE $Qp^{(k+1)} = \bar{p}^{(k+1)}$

The cost per iteration is two matrix multiplies (by A and by A^t), two solves of systems of the form $Mz = d$, and $5n$ multiplies, which is essentially the cost of two GCG iterations.

The iterate $y^{(k)}$ can be characterized as the unique element in the affine Krylov subspace

$$x^{(0)} + (Q^{-1}A^t)\,\mathrm{Span}\{r^{(0)}, (Q^{-1}AQ^{-1}A^t)r^{(0)}, \ldots, (Q^{-1}AQ^{-1}A^t)^{k-1}r^{(0)}\} \equiv x^{(0)}+\mathcal{T}_m$$

satisfying the orthogonality condition

$$(z, y^{(k)}-x)_Q = 0 \qquad \text{for all } z \in \mathcal{T}_k.$$

Thus

$$y^{(k)} = \underset{y \in x^{(0)}+\mathcal{T}_k}{\arg\min} \|y-x\|_Q \ ,$$

and the standard analysis based on the Chebyshev polynomials yields the error bound

$$\|y^{(k)}-x\|_Q \le \frac{2}{\rho^k + \rho^{-k}} \|x^{(0)}-x\|_Q \ , \tag{5}$$

where $\rho = (\kappa+1)/(\kappa-1)$ and κ is the condition number of $\hat{A}$.

If $Q = M$, then

$$x^{(0)}+\mathcal{T}_k = x^{(0)} + (I+K)\,\mathrm{Span}\{v^{(0)}, K^2v^{(0)}, \ldots, K^{2k-2}v^{(0)}\} \subseteq x^{(0)}+(I+K)S_{2k}$$

and

$$y^{(k)} = \underset{y \in x^{(0)}+\mathcal{T}_k}{\arg\min} \|y-x\|_M \ . \tag{6}$$

Moreover, since[3] $\kappa = \sqrt{1+A^2}$, the error bound (5) reduces to (3).

But an even stronger relationship exists between GCG and CSP. If m is even ($= 2k$), then $\pi_m(\mu)$ is even and $\pi_m(1) = 1$ so that

$$\pi_m(\mu) = 1 - (1+\mu)\,p_{2k-2}(\mu)\,(1-\mu) \ ,$$

where $p_{2k-2}(\mu)$ is an even polynomial of degree at most $2k-2$. Thus, by equation (2),

---

[3] If the number of equations is even, then $\sqrt{1+A^2}$ may only be an upper bound on κ.

$$x^{(2k)} = x^{(0)} - (I+K)p_{2k-2}(K)(I-K)(x^{(0)}-x)$$

$$= x^{(0)} + (I+K)p_{2k-2}(K)v^{(0)}$$

$$\in x^{(0)}+\mathcal{T}_k \ ;$$

i.e., $x^{(2k)}$, the best approximation to x from the affine subspace $x^{(0)}+(I+K)S_{2k}$, lies in the smaller affine subspace $x^{(0)}+\mathcal{T}_k$. But since $y^{(k)}$ is the best approximation to x from $x^{(0)}+\mathcal{T}_k$ (see (6)), it follows that $x^{(2k)} = y^{(k)}$. Hageman, Luk, and Young [8] and Elman [5] give different proofs that the two methods are "virtually equivalent."

The cost of computing $y^{(k)}$ is essentially the same as the cost of computing $x^{(2k)}$. However, the odd iterates generated by GCG could be better approximations to x than the even iterates (although by at most a constant factor in view of (4)). Since, in addition, GCG requires somewhat less storage, it is probably the better method.

4. Two-Level Methods

But what if systems of the form $Mz = d$ are *not* easy to solve? Golub and Overton [7] have proposed a modification[4] of GCG in which the step

$$\text{SOLVE} \quad Mv^{(m)} = r^{(m)}$$

is replaced by

$$\text{FIND SOME } v^{(m)} \text{ SATISFYING } \|Mv^{(m)} - r^{(m)}\|_M \leq \delta \|r^{(m)}\|_M \ ,$$

where $0 \leq \delta < 1$ is some constant.[5] This is implemented using an *inner* iterative method to find $v^{(m)}$ on the m^{th} *outer* iteration. Basing the stopping criterion on the size of the relative residual has the effect of solving $Mv^{(m)} = r^{(m)}$ to increasing absolute accuracy as $x^{(m)}$ converges to x.

While they were unable to analyze this two-level scheme, Golub and Overton [7] did analyze a similar scheme using the two-stage Richardson method (also a three-term recurrence) as the outer iteration. As one would expect, taking δ closer to 0 results in a larger number of inner iterations per outer iteration and a smaller number of outer iterations; whereas taking δ closer to 1 results in a smaller number of inner iterations per outer iteration but a larger number of outer iterations. The same behavior for the two-level GCG method can be seen in the numerical

---

[4] Dembo, Eisenstat, and Steihaug [3] analyze a similar modification to Newton's method for nonlinear systems of equations.

[5] The case $\delta = 0$ corresponds to the original GCG method.

results presented in Section 5.

One could take a similar approach with the CSP method. Since M is symmetric and positive definite, a logical choice for the inner iteration would be the preconditioned conjugate gradient method with some preconditioning matrix $\bar{Q}$ (see [1]). But why use a two-level iteration at all when one can simply take $Q = \bar{Q}$ (instead of $Q = M$)? The numerical results presented in Section 5 suggest that this approach is superior.

5. Numerical Results

In this section, we reproduce the numerical experiments reported by Golub and Overton [7] for the two-level GCG method and present the corresponding results for the CSP method.

Consider the elliptic partial differential equation

$$-\Delta u + (au)_x + au_x + (bu)_y + bu_y + cu = f$$

subject to Dirichlet boundary conditions on the unit square $[0,1]\times[0,1]$, where

$$a(x,y) = 5e^{x^2+y^2}, \quad b(x,y) = 5e^{x^2+y^2}, \quad c(x,y) = 10e^{2(x^2+y^2)}$$

and $f(x,y)$ and the boundary conditions are chosen to make the solution $u(x,y) = e^{x^2+y^2}$.

The five-point centered finite-difference discretization on a rectilinear grid with n interior mesh points in each direction leads to a system of n^2 linear equations in which Mu corresponds to $-\Delta u + cu$. Thus we use the fast Poisson solver HWSCRT from FISHPACK [9] as a preconditioning for CSP and for an inner preconditioned conjugate gradient iteration in the two-level GCG scheme. In each case, the stopping criterion was $\|r^{(m)}\| \leq 10^{-8}$.

The numbers of (outer) iterations and Poisson solves are given in Table 1 and the number of Poisson solves is plotted against δ in Figure 1. Clearly CSP is a better method than GCG for this problem, even with the optimal choice of δ.

6. Conclusions

As we have seen, if systems of the form $Mz = d$ are "easy" to solve, then GCG is better than CSP. If not, then CSP is superior. Of course, it is not clear that either method is the best possible for this class of problems.

References

[1] Rati Chandra. *Conjugate Gradient Methods for Partial Differential Equations.* Ph.D. Thesis, Department of Computer Science, Yale University, 1978. Also available as Research Report #129.

[2] Paul Concus and Gene H. Golub. A generalized conjugate gradient method for nonsymmetric systems of linear equations. In R. Glowinski and J. L. Lions, Editors, *Lecture Notes in Economics and Mathematical Systems, Volume 134*, Springer Verlag, 1976, pp. 56-65.

[3] R. S. Dembo, S. C. Eisenstat, and T. Steihaug. Inexact Newton methods. *SIAM Journal on Numerical Analysis* 19(2):400-408 (1982).

[4] Stanley C. Eisenstat. *A Note on the Generalized Conjugate Gradient Method.* Technical Report #228, Department of Computer Science, Yale University, 1982. To appear in *SIAM Journal on Numerical Analysis.*

[5] Howard C. Elman. *Iterative Methods for Large, Sparse, Nonsymmetric Systems of Linear Equations.* Ph.D. Thesis, Department of Computer Science, Yale University, 1982. Also available as Technical Report #229.

[6] D. K. Faddeev and V. N. Faddeeva. *Computational Methods of Linear Algebra.* W. H. Freeman and Company, San Francisco, 1963.

[7] Gene H. Golub and Michael L. Overton. Convergence of a two-stage Richardson iterative procedure for solving systems of linear equations. In G. A. Watson, Editor, *Lecture Notes in Mathematics, Volume 912*, Springer Verlag, 1982, pp. 125-139. Proceedings of the 9th Biennial Conference, Dundee, Scotland, June 23-26, 1981.

[8] L. A. Hageman, Franklin T. Luk, and David M. Young. On the equivalence of certain iterative acceleration methods. *SIAM Journal on Numerical Analysis* 17:852-873 (1980).

[9] P. Swarztrauber and R. Sweet. *Efficient FORTRAN Subprograms For the Solution of Elliptic Equations.* Technical Report TN/IA-109, NCAR, July 1975.

[10] Olof Widlund. A Lanczos method for a class of nonsymmetric systems of linear equations. *SIAM Journal on Numerical Analysis* 15:801-812 (1978).

Table 1: Number of (Outer) Iterations and Poisson Solves

		$n = 15$		$n = 31$	
	δ	m	m_p	m	m_p
GCG:	.0	33		36	
	.00100	36	266	39	279
	.00158	36	250	39	270
	.00251	36	247	39	261
	.00398	37	238	41	255
	.00631	38	226	41	244
	.0100	38	218	42	235
	.0158	39	209	42	216
	.0251	40	198	43	209
	.0398	41	193	45	196
	.0631	44	175	47	181
	.100	52	182	71	237
	.158	135	398	273	761
CSP:		39	78	42	84

m = number of (outer) iterations
m_p = number of Poisson solves

Figure 1: Number of Poisson Solves vs. δ